LED: Dynamic Epistemic Logic with Neutrosophic Semantics

for Language Model Hallucinations

A Position Paper and Research Agenda

Maikel Yelandi Leyva Vázquez
Universidad de Guayayquil
Universidad Bolivariana del Ecuador
Florentin Smarandache
University of New Mexico, Department of Mathematics

ABSTRACT

Large language models (LLMs) hallucinate with confidence despite uncertainty. We argue this stems from a fundamental absence: LLMs lack formal machinery to represent epistemic indeterminacy—the genuine uncertainty when evidence is contradictory, incomplete, or contested. We propose LED (Dynamic Epistemic Logic with Neutrosophic Semantics), a framework for representing evolving epistemic states $\sigma(t) = (T(t), I(t), F(t)) \in [0,1]^3$ through three operators: Refinement, Conflict Detection, and Resolution. LED detects three hallucination patterns: Unjustified Confidence, Unresolved Conflict, and Indeterminacy Divergence. This position paper articulates the theoretical foundation and invites the community to validate and extend LED.

Keywords: epistemic logic, neutrosophic sets, language models, hallucinations, uncertainty representation, machine reasoning

1. THE PROBLEM: HALLUCINATIONS AS EPISTEMIC FAILURES

1.1 Why LLMs Hallucinate

Large language models trained via next-token prediction achieve remarkable fluency—but at a cost. The training objective (minimize cross-entropy over tokens) is fundamentally misaligned with the epistemic requirements of truthful reasoning.

Consider three scenarios:

Scenario A: Established Facts (Healthcare)

Question: "What is the standard adult dose of acetaminophen?"
LLM: "4,000 mg per day for short-term use."
Reality: Well-established (FDA guidelines). Evidence is overwhelming and uncontested.
Model behavior: Confident, appropriate.

Scenario B: Debated Questions (Policy)

Question: "Was Moore v. Harper the most important SCOTUS decision in 2023?"
LLM: "Yes, it established crucial limits on state election power."
Reality: Moore v. Harper WAS important, but other decisions rival it. Experts disagree.
Model behavior: Confident assertion masks genuine disagreement.

Scenario C: Open Questions (Emerging Tech)

Question: "Can current LLMs perform genuine logical reasoning?"
LLM: "Chain-of-thought prompting demonstrates reasoning capabilities."
Reality: This is actively debated in academia and AI research communities.
Model behavior: Commits to one side of an open scientific question as if settled.

The Pattern: Hallucinations follow a structure. The model asserts with high confidence in situations where uncertainty is justified. The problem is not random failure—it is systematic misrepresentation of epistemic state.

1.2 Why Existing Approaches Fail

Classical Confidence Scores A single scalar (e.g., P(statement|evidence) = 0.85) cannot distinguish justified from unjustified high confidence. A confidence of 0.85 tells you nothing about why the model is confident or whether confidence is appropriate.

Ensemble Methods & Uncertainty via Disagreement When models disagree, does this reflect genuine epistemic uncertainty or just model variance? The method conflates these two very different phenomena.

Semantic Consistency & Embedding Similarity Measuring internal consistency tells you the model is coherent, not that it is correct or that uncertainty is justified.

2. THE THEORETICAL FOUNDATION

2.1 Why Epistemic Logic?

Epistemic logic formalizes how agents know and believe. An agent's epistemic state captures: what the agent knows, what the agent believes, what the agent treats as uncertain, and how beliefs evolve with new information.

Key insight for LLMs: LLMs do not have an explicit epistemic state. They have parameters that encode statistical patterns, but no formal representation of uncertainty, contradiction, or ignorance.

Dynamic Epistemic Logic (DEL) extends classical epistemic logic by modeling how states evolve when information arrives. We adapt DEL to LLMs by replacing classical possible-worlds semantics with neutrosophic valuations.

2.2 Why Neutrosophic Logic?

Standard logic admits two truth values (True/False). Probabilistic approaches use scalar confidence [0,1]. But the epistemology of hallucinations requires three independent dimensions:

- **T (Truth):** How much evidence supports the proposition?
- **I (Indeterminacy):** How much genuine uncertainty (contradiction, disagreement, unknowns) exists?
- **F (Falsity):** How much evidence contradicts the proposition?

Why independence matters: In classical logic, ¬T = F (if not true, then false). But in genuine uncertainty: T and F can both be high (conflicting evidence), I can be high even when T is moderate (disagreement about interpretation), and multiple values can be low (genuine gaps in knowledge).

Statement	T	I	F	Interpretation
Water boils at 100°C	0.95	0.05	0.00	Established fact
Moore v. Harper was most important	0.60	0.35	0.30	Conflicting evidence
LLMs can reason	0.40	0.45	0.35	Genuinely debated

Neutrosophic sets (Smarandache, 1998) provide exactly this formalism. Used in engineering (image denoising, medical diagnosis), but underexplored in LLM semantics.

2.3 Formal Framework: LED

Definition: An epistemic state is $\sigma = (T, I, F) \in [0,1]^3$

Confidence function: $C(\sigma) = \max(0, T - I - F)$ (from Wang et al., 2010 SVNS algebra)

Decision Rules:

$C > 0.60$: High justified confidence → ANSWER

$0.30 < C < 0.60$: Moderate confidence → ANSWER WITH CAVEATS

$C < 0.30$ or $I > 0.40$: Low confidence or high indeterminacy → ABSTAIN

3. THREE HALLUCINATION PATTERNS

LED formalizes three ways hallucinations manifest:

Pattern 1: Unjustified Confidence

$$P_1(\sigma) \leftrightarrow (C(\sigma) > 0.30) \wedge (I(\sigma) > 0.40)$$

Model asserts with measurable confidence despite very high indeterminacy.

Example: Question: "Is X the best economic policy?" (inherently debated)

Model: "Yes, because..." (confident assertion)

LED detects: C = 0.35, I = 0.50 → **Pattern 1 triggered** → Recommend ABSTAIN

Pattern 2: Unresolved Conflict

$$P_2(seq) \leftrightarrow \exists t: (T > 0.30 \wedge F > 0.30) \wedge \neg Resolved$$

Model has evidence on both sides but doesn't acknowledge or reconcile the conflict.

Example: Question: "Most important SCOTUS decision 2023?"

Analysis: T = 0.75 (evidence for), F = 0.35 (evidence against) → **Pattern 2 triggered**

Pattern 3: Indeterminacy Divergence

$$P_3(seq) \leftrightarrow \exists t: I(\sigma_t) > I(\sigma_{t-1}) + \delta \text{ where } \delta = 0.05$$

As reasoning progresses, indeterminacy INCREASES instead of DECREASING. This signals the model is discovering complexity, not resolving it.

4. WHY THIS MATTERS

4.1 Conceptual Significance

Hallucinations are not "random errors." They are failures of epistemic representation—the model lacks machinery to distinguish "I know this with certainty" from "I'm pattern-matching but uncertain." LED makes this distinction explicit.

4.2 Methodological Significance

Classical confidence (probability, softmax) works well for classification. But truthfulness requires epistemic reasoning. Neutrosophic logic, underused in modern ML, is precisely the right mathematical tool for this setting.

4.3 Practical Significance

LLMs are deployed in medical diagnosis, legal analysis, and policy—domains where high-confidence hallucinations cause real harm. A framework that forces LLMs to represent uncertainty explicitly is foundational for responsible AI.

5. RESEARCH AGENDA

We identify five priority research directions:

5.1 Empirical Validation

Do the three hallucination patterns predict actual LLM failures? Benchmark LED on 300+ questions across 5+ domains with expert annotation.

5.2 Multi-Model Generalization

Is LED universal or model-specific? Validate on GPT-4, Llama, Gemini, and open-source models.

5.3 Mechanistic Interpretability

Why does the model assign T=0.75 vs T=0.70? Use attention visualization, probing tasks, and layer-wise analysis.

5.4 Integration with RAG

How should LED interact with Retrieval-Augmented Generation systems? Extend to scenarios where retrieval + generation co-determine (T, I, F).

5.5 Deployment & User Studies

Do users trust systems that abstain when uncertain? Is abstention preferable to confident hallucination in real settings? Deploy in medical QA, legal research, policy analysis.

6. LIMITATIONS AND OPEN QUESTIONS

6.1 Representation vs Reality

(T, I, F) represent the model's epistemic state derived from reasoning, not ground truth. A well-reasoned hallucination could have high T but be factually wrong.

6.2 Threshold Selection

LED uses thresholds (C > 0.30, I > 0.40). Are these universal? Domain-specific? How should they be set?

6.3 Explainability of Assignments

We can explain the framework (what T, I, F mean). We cannot explain why a specific question gets T=0.75 vs T=0.80. This gap is critical for high-stakes applications.

7. CALL TO ACTION

This position paper argues:

- **LLM hallucinations are epistemic failures,** not random noise
- **Neutrosophic logic provides the right formalism** for representing indeterminacy
- **Dynamic Epistemic Logic enables reasoning about evolving states** as new information arrives
- **Three formal patterns characterize hallucinations** (Unjustified Confidence, Unresolved Conflict, Indeterminacy Divergence)

We invite the community to:

- Validate LED on your own data (medical QA, legal analysis, technical documentation)
- Extend LED to new domains, multi-model settings, mechanistic interpretability
- Deploy LED in production systems and report real-world results
- Contribute theory (refined thresholds, additional operators, formal proofs)
- Discuss limitations (what doesn't work, where LED fails, why)

8. CONCLUSION

LLMs will be more trustworthy when they can say "I don't know" with the same confidence they currently say "I know." LED is a framework for formalizing this capability.

The stakes are high. As LLMs move into medical, legal, and policy domains, the ability to represent and manage epistemic uncertainty is not a nice-to-have—it is essential.

We release this as a position paper to catalyze community discussion, validation, and extension. The framework is ready for empirical stress-testing. We welcome collaborators.

REFERENCES

[1] Hintikka, J. (1962). Knowledge and Belief. Cornell University Press.

[2] van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). Dynamic Epistemic Logic. Springer.

[3] Smarandache, F. (1998). Neutrosophy: Neutrosophic Probability, Set, and Logic. American Research Press.

[4] Wang, H., Smarandache, F., Zhang, Y., & Sunderraman, R. (2010). Single Valued Neutrosophic Sets. Multispace and Multistructure.

[5] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. ICML.

[6] Anthropic. (2025). Constitutional AI and LLM Behavior. Technical Report.

AUTHOR INFORMATION

Maikel Yelandi Leyva Vázquez

Universidad de Guayaquil
Universidad Bolivariana del Ecuador

Email: maikel.leyva@ube.edu.ec
ORCID: 0000-0002-9486-5093

Florentin Smarandache

University of New Mexico
Department of Mathematics
Email: smarand@unm.edu
ORCID: 0000-0002-5560-5926

Status: Position Paper - Open for Community Feedback
Preprint: ResearchGate, 2026